

#4

00862.022404

PATENT APPLICATION



IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of:

Kenichi FUJII, et al.

Application No.: 09/972,996

Filed: October 10, 2001

For: SPEECH RECOGNITION SYSTEM,
SPEECH RECOGNITION
APPARATUS, AND SPEECH
RECOGNITION METHOD

)
:
Examiner: Unassigned
)
:
Group Art Unit: 2644
)
:
)
:
January 3, 2002
:
)
:
)
:
:

Commissioner for Patents
Washington, D.C. 20231

SUBMISSION OF PRIORITY DOCUMENT (DOCUMENTS)

Sir:

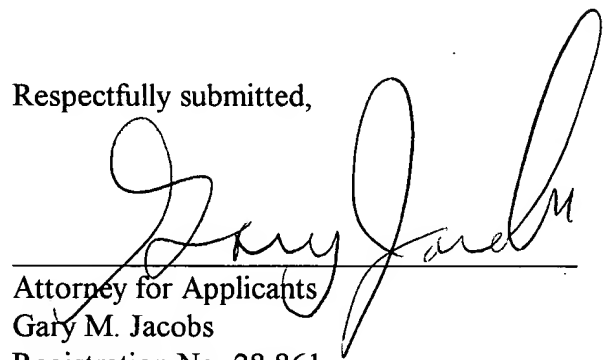
In support of Applicants' claim for priority under 35 U.S.C. § 119, enclosed are
certified copies of the following foreign applications:

2000-311098 filed October 11, 2000;

2000-378019, filed December 12, 2000.

Applicants' undersigned attorney may be reached in our Washington, D.C. office by telephone at (202) 530-1010. All correspondence should continue to be directed to our address given below.

Respectfully submitted,



Attorney for Applicants
Gary M. Jacobs
Registration No. 28,861

FITZPATRICK, CELLA, HARPER & SCINTO
30 Rockefeller Plaza
New York, New York 10112-3801
Facsimile: (212) 218-2200

GMJ/dc

DC_MAIN 82786 v 1

(translation of the front page of the priority document of
Japanese Patent Application No. 2000-311098)



PATENT OFFICE
JAPANESE GOVERNMENT

This is to certify that the annexed is a true copy of the
following application as filed with this Office.

Date of Application: October 11, 2000

Application Number : Patent Application 2000-311098

Applicant(s) : Canon Kabushiki Kaisha

November 2, 2001

Commissioner,
Patent Office

Kouzo OIKAWA

Certification Number 2001-3096813

CFM 2404 US



日本国特許庁
JAPAN PATENT OFFICE

09/972,996
Kenichi Fujii
October 10, 2001

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office

出願年月日
Date of Application: 2000年10月11日

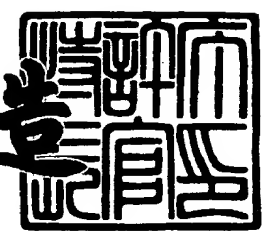
出願番号
Application Number: 特願2000-311098

出願人
Applicant(s): キヤノン株式会社

2001年11月 2日

特許庁長官
Commissioner,
Japan Patent Office

及川耕造



出証番号 出証特2001-3096813

【書類名】 特許願

【整理番号】 4153123

【提出日】 平成12年10月11日

【あて先】 特許庁長官殿

【国際特許分類】 H04N 5/00

【発明の名称】 音声処理装置、音声処理方法及び記憶媒体

【請求項の数】 5

【発明者】

 【住所又は居所】 東京都大田区下丸子3丁目30番2号 キヤノン株式会社
社内

 【氏名】 藤井 憲一

【発明者】

 【住所又は居所】 東京都大田区下丸子3丁目30番2号 キヤノン株式会社
社内

 【氏名】 伊藤 史朗

【発明者】

 【住所又は居所】 東京都大田区下丸子3丁目30番2号 キヤノン株式会社
社内

 【氏名】 上田 隆也

【発明者】

 【住所又は居所】 東京都大田区下丸子3丁目30番2号 キヤノン株式会社
社内

 【氏名】 池田 裕治

【特許出願人】

 【識別番号】 000001007

 【氏名又は名称】 キヤノン株式会社

【代理人】

 【識別番号】 100090273

 【弁理士】

【氏名又は名称】 國分 孝悦

【電話番号】 03-3590-8901

【手数料の表示】

【予納台帳番号】 035493

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 9705348

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 音声処理装置、音声処理方法及び記憶媒体

【特許請求の範囲】

【請求項 1】 音声を入力する音声入力手段と、

ネットワークに接続された複数の音声認識装置の中の一つを指定する指定情報を前記音声入力手段に入力された音声から検出する検出手段と、

前記検出手段で検出された指定情報に対応する音声認識装置を用いて前記音声入力手段に入力された音声を音声認識する音声認識手段とを有することを特徴とする音声処理装置。

【請求項 2】 前記音声認識手段は更に、前記指定情報が検出されなかった場合には、前記複数の音声認識装置を用いることなく前記音声入力手段に入力された音声を音声認識することを特徴とする請求項 1 に記載の音声処理装置。

【請求項 3】 音声を入力する音声入力工程と、

ネットワークに接続された複数の音声認識装置の中の一つを指定する指定情報を前記音声入力工程で入力された音声から検出する検出工程と、

前記検出工程で検出された指定情報に対応する音声認識装置を用いて前記音声入力工程で入力された音声を音声認識する音声認識工程とを有することを特徴とする音声処理方法。

【請求項 4】 前記音声認識工程は更に、前記指定情報が検出されなかった場合には、前記複数の音声認識装置を用いることなく前記音声入力工程で入力された音声を音声認識することを特徴とする請求項 3 に記載の音声処理方法。

【請求項 5】 請求項 3 または 4 に記載の音声処理方法を実現するためのプログラムを記憶したこと特徴とする記憶媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、ネットワークに接続された複数の音声認識装置を利用する音声処理装置、音声処理方法及び記憶媒体に関する。

【0002】

【従来の技術】

近年、人間が発声した音声を所定の規則に従って電子計算機上で認識する技術（所謂、音声認識技術）が実用化されている。更に、近年では、この音声認識技術を用いて、従来手動で電子計算機に入力していたコマンドや文字情報を音声で入力する技術が開発されている。

【0003】

【発明が解決しようとする課題】

しかしながら、音声認識は比較的計算量の多い処理であるので、ユーザが発声した全ての音声を実時間で認識するためには高価で高性能な電子計算機が必要となる。そのため、このような音声認識をモバイルコンピュータや携帯電話等の小型で安価な携帯端末装置に適用することは困難であるという問題があった。

【0004】

本発明は前述の問題に鑑みてなされたものであり、ネットワークに接続された複数の音声認識装置を利用して高精度な音声認識を効率的に行うことを目的とする。

【0005】

【課題を解決するための手段】

本発明の音声処理装置は、音声を入力する音声入力手段と、ネットワークに接続された複数の音声認識装置の中の一つを指定する指定情報を前記音声入力手段に入力された音声から検出する検出手段と、前記検出手段で検出された指定情報に対応する音声認識装置を用いて前記音声入力手段に入力された音声を音声認識する音声認識手段とを有することを特徴とする。

【0006】

また、本発明の音声処理方法は、音声を入力する音声入力工程と、ネットワークに接続された複数の音声認識装置の中の一つを指定する指定情報を前記音声入力工程で入力された音声から検出する検出工程と、前記検出工程で検出された指定情報に対応する音声認識装置を用いて前記音声入力工程で入力された音声を音声認識する音声認識工程とを有することを特徴とする。

【0007】

また、本発明の記憶媒体は、上記音声処理方法を実現するためのプログラムを記憶したことを特徴とする。

【0008】

【発明の実施の形態】

以下、図面を参照して本発明の実施の形態を詳細に説明する。

図1は、本実施の形態に係る音声認識システムの基本構成を示す図である。

【0009】

図1において、101はモバイルコンピュータや携帯電話等の通信端末装置であり、小語彙辞書を持つ音声認識プログラムを内蔵する。102及び103は大語彙辞書を持つ高性能な音声認識装置であり、それぞれ採用する文法規則が異なる。104はインターネット、移動体通信網等のネットワークである。

【0010】

通信端末装置101は、演算量が少なく安価で簡易な音声認識装置であり、主に「戻る（モドル）」、「進む（ススム）」等の単純で短い単語を簡易に音声認識する機能を有する。これに対して、音声認識装置102及び103は、演算量が多く高価で高精度な音声認識装置であり、主に住所、氏名等の複雑で長い連続文を高精度に音声認識する機能を有する。このように、本実施の形態の音声認識システムでは、音声認識の機能を分散化することによって、ユーザに提供する情報端末装置の認識効率を落とすことなく安価に構成し、利便性や携帯性を高めている。

【0011】

通信端末装置101と音声認識装置102及び103とは、ネットワーク104を介してデータ通信可能である。通信端末装置101に入力されたユーザの音声は、ユーザがキーワードで指定する音声認識装置102又は103に伝送される。本実施例では、音声認識装置102を指定するキーワードを「入力1（ニュウリョクイチ）」とし、音声認識装置103を指定するキーワードを「入力2（ニュウリョクニ）」とする。音声認識装置102又は103は、通信端末装置101からの音声（キーワードの部分を除く）を音声認識し、音声認識によって得られた文字列を通信端末装置101に返信する。

【0012】

次に、図2を用いて、本実施の形態に係る通信端末装置101の構成を説明する。

【0013】

図2において、201は制御部、202は記憶部、203は通信部、204は音声入力部、205は操作部、206は音声出力部、207は表示部である。また、208はアプリケーションプログラム、209は音声認識プログラム、210はユーザインタフェース制御プログラム、211はキーワード登録部である。

【0014】

制御部201は、ワークメモリやマイクロコンピュータ等から構成され、記憶部202に記憶されたアプリケーションプログラム208、音声認識プログラム209及びユーザインタフェース制御プログラム210を読み出して実行する。

【0015】

記憶部202は、磁気ディスク、光ディスク、ハードディスク装置等の記憶媒体から構成され、アプリケーションプログラム208、音声認識プログラム209、ユーザインタフェース制御プログラム210及びキーワード登録部211を所定の領域に記憶する。通信部203は、ネットワーク104に接続された音声認識装置102及び103とデータ通信を行う。

【0016】

音声入力部204は、マイクロフォン等から構成され、ユーザが発声した音声を入力する。操作部205は、キーボード、マウス、タッチパネル、ジョイスティック、ペン、タブレット等から構成され、アプリケーションプログラム208のグラフィカルユーザインタフェースを操作する。

【0017】

音声出力部206は、スピーカやヘッドフォン等から構成される。表示部207は、液晶ディスプレイ等の表示器から構成され、アプリケーションプログラム208のグラフィカルユーザインタフェースを表示する。

【0018】

アプリケーションプログラム208は、ネットワーク104上の情報（ホーム

ページや各種のデータファイル等のウェブコンテンツ)を閲覧するためのウェブブラウザの機能とこの機能进行操作するためのグラフィカルユーザインタフェースを有する。音声認識プログラム209は、主に「戻る(モドル)」、「進む(ススム)」等の単純で短い単語を簡易に音声認識する機能を有するプログラムである。ユーザインタフェース制御プログラム210は、音声認識プログラム209を用いて音声認識して得た文字列を所定のコマンドに変換してアプリケーションプログラム208に入力したり、音声認識装置102又は103を用いて音声認識して得た文字列をアプリケーションプログラム208に入力したりする。キーワード登録部211は、ネットワーク104に接続された音声認識装置102及び103を指定するためのキーワードを登録する。

【0019】

次に、図3を参照しながら、本実施の形態に係る通信端末装置101が音声認識装置102及び103を指定するキーワードを登録する手順を説明する。この手順は、制御部201が記憶部202に記憶されたユーザインタフェース制御プログラム210に従って実行する。

【0020】

ステップS301において、制御部201は、表示部207を用いてキーワードの登録されていない音声認識装置をユーザに通知する。ユーザは、操作部205を用いて音声認識装置102又は103を指定するキーワードを入力する。本実施例では、音声認識装置102を指定するキーワードを「入力1(ニユウリョクイチ)」とし、音声認識装置103を指定するキーワードを「入力2(ニユウリョクニ)」とする。

【0021】

ステップS302において、制御部201は、音声認識装置102又は103を指定するキーワードをキーワード登録部211に登録する。ステップS303において、制御部201は、音声認識装置102及び103のキーワードを登録したか否かを判定する。全ての音声認識装置のキーワードを登録した場合には登録の処理を終了する。

【0022】

次に、図4を参照しながら、本実施の形態に係る通信端末装置101が、ネットワーク104に接続された音声認識装置102又は103を用いて、入力音声を実声認識する手順を説明する。この手順は、制御部201が記憶部202に記憶されたユーザインタフェース制御プログラム210に従って実行する。

【0023】

ステップS401において、制御部201は、音声入力部204に入力されたユーザの音声を音声認識プログラム209に入力する。本実施の形態では、外部の音声認識装置102又は103を用いて音声認識を行う場合には、認識対象となる文字列を発声する前に必ずキーワードを発声するようにする。例えば、音声認識装置102を用いて音声認識を行う場合、ユーザは「ニューヨークイチ（ポーズ）カワサキシ」と発声する。このように構成することにより、ユーザとしてはどの音声認識装置を使用したいかを意識して選択することができ、通信端末装置101にとってはキーワードの検出が容易となり、検出処理の高速化を図ることができる。

【0024】

ステップS402において、制御部201は音声認識プログラム209を用いてステップS401で入力された音声を簡易に音声認識し、認識した文字列からキーワード登録部211に登録されたキーワードを検出する。

【0025】

ステップS403において、制御部201は、ステップS403でキーワードが検出された否かを判定する。検出された場合にはステップS404に進み、検出されなかった場合にはステップS407に進む。例えば、ユーザが「ニューヨークイチ（ポーズ）カワサキシナカハラクイマイカミマチキヤノンコスギジギョウショ」と発声した場合には、音声認識装置102を指定するキーワード「ニューヨークイチ」が検出されるので、ステップS404に進む。また例えば、ユーザが「モデル」や「ススム」と発声した場合には、キーワード登録部211に登録されたキーワードは検出されないで、ステップS407に進む。

【0026】

ステップS404において、制御部201は、ステップS402で検出された

キーワードに対応する音声認識装置102又は103を選択する。つまり、通信端末装置101は、キーワード登録部211に登録されたキーワードが検出された場合には、ネットワーク104に接続された複数の音声認識装置の中の一つを選択して音声認識を行う。従って、ユーザが「ニュウリョクイチ（ポーズ）カワサキシナカハラクイマイカミマチキヤノンコスギジギョウショ」と発声した場合には、音声認識装置102を選択する。

【0027】

ステップS405において、制御部201は、ステップS403で選択された音声認識装置102又は103にステップS401で入力された音声（但し、キーワードの部分を除く）を送信する。このように、キーワードの部分を除いた音声を送信することにより、通信効率を高め、音声認識装置102及び103が不要な音声認識を行うことを防止する。音声認識装置102又は103は、通信端末装置101からの音声を音声認識し、認識した文字列を通信端末装置101に返信する。ユーザが「ニュウリョクイチ（ポーズ）カワサキシナカハラクイマイカミマチキヤノンコスギジギョウショ」と発声した場合、音声認識装置102は、文字列「川崎市中原区今井上町キヤノン小杉事業所」を音声認識し、この文字列を通信端末装置101に返信する。

【0028】

ステップS406において、制御部201は、音声認識装置102又は103で認識された文字列をアプリケーションプログラム208に入力する。アプリケーションプログラム208は、入力された文字列を表示部207に表示されたグラフィカルユーザインタフェース上の予め選択された入力フィールドに出力する。

【0029】

一方、ステップS407において、制御部201は、音声認識プログラム209を用いてステップS401で入力された音声の音声認識を行う。つまり、通信端末装置101は、キーワード登録部211に登録されたキーワードが検出されなかった場合には、内部の音声認識プログラム209を用いて自動的に音声認識を行う。従って、ユーザが「モドル」や「ススム」と発声した場合には、キーワ

ードは検出されないので音声認識プログラム209を用いて自動的に音声認識し、文字列「戻る」や「進む」を得る。

【0030】

ステップS408において、制御部201は、音声認識プログラム209で認識された文字列を所定のコマンドに変換し、変換されたコマンドをアプリケーションプログラム208に入力する。例えば、文字列「戻る」は、現在閲覧中のページを一つ前に戻すという意味のコマンドに変換され、文字列「進む」は、現在閲覧中のページを一つ前に進めるという意味のコマンドに変換される。アプリケーションプログラム208は、入力されたコマンドに対応する処理を実行し、実行された結果を表示部207に表示する。

【0031】

以上説明したように、本実施の形態によれば、演算量が少なく安価で簡易な音声認識についてはユーザに提供する通信端末装置で実行し、演算量が多く高価で高精度な音声認識についてはネットワークに接続された複数の音声認識装置の何れかで実行するので、ユーザに提供する通信端末装置を、認識効率を落とすことなく安価に構成することができる。

【0032】

また、本実施の形態によれば、ネットワークに接続された複数の高精度な音声認識装置の一つをユーザが発声するキーワードによって指定できるので、手による煩雑な操作を削減することができる。また、専用の操作ボタン等を設けなくてもよくなるため、ユーザに提供する通信端末装置を小型に構成することができる。特に、モバイルコンピュータや携帯電話等の携帯端末装置では、利便性や携帯性を高めることができる。

【0033】

また、本実施の形態によれば、入力音声にキーワードが含まれているか否かに応じて入力音声の内部の音声認識プログラムで認識するか、外部の音声認識装置で認識するかを簡単に判別することができる。

【0034】

本実施の形態では、ネットワーク104に接続された2つの音声認識装置10

2 及び 1 0 3 を用いて音声認識システムを構成する例について説明したが、本発明はこのような構成に限るものではない。2 つ以上の音声認識装置を用いて音声認識システムを構成することも可能である。この場合、ユーザは各音声認識装置を指定するキーワードをキーワード登録部 2 1 1 に登録する。これらの音声認識装置を利用する場合にはキーワード登録部 2 1 1 に登録されたキーワードを発声する。また、複数の異なる音声認識ユニットを有する音声認識装置を用いて音声認識システムを構成することも可能である。この場合、ユーザは 1 つの装置が有する複数の異なる音声認識ユニットの夫々を指定するキーワードをキーワード登録部 2 1 1 に登録する。これらの音声認識ユニットを利用する場合にはキーワード登録部 2 1 1 に登録されたキーワードを発声する。

【 0 0 3 5 】

（その他の実施の形態）

本発明は、上記実施の形態に限るものではなく、様々な形態で実施することができる。

【 0 0 3 6 】

例えば、制御部 2 0 1 が読み出したユーザインタフェース制御プログラム 2 1 0 の指示に基づき、制御部 2 0 1 上で稼働している OS（オペレーティングシステム）などが上記実施の形態の処理の一部または全部を行い、その処理によって上記実施の形態が実現される場合にも本発明は適用可能である。

【 0 0 3 7 】

また、記憶部 2 0 2 から読み出されたユーザインタフェース制御プログラム 2 1 0 が、情報端末装置 1 0 1 に接続された機能拡張ユニットに備わるメモリに書き込まれた後、そのプログラム 2 1 0 の指示に基づいてその機能拡張ユニットに備わる制御部などが実際の処理の一部または全部を行い、その処理によって上記実施の形態が実現される場合にも本発明は適用可能である。

【 0 0 3 8 】

【発明の効果】

以上説明したように、本発明によれば、ネットワークに接続された複数の音声認識装置を利用して高精度な音声認識を効率的に行うことができる。

【図面の簡単な説明】

【図 1】

本実施の形態に係る音声認識システムの構成を示す図である。

【図 2】

本実施の形態に係る通信端末装置の構成を示すブロック図である。

【図 3】

本実施の形態に係る通信端末装置がキーワードを登録する手順を説明するフローチャートである。

【図 4】

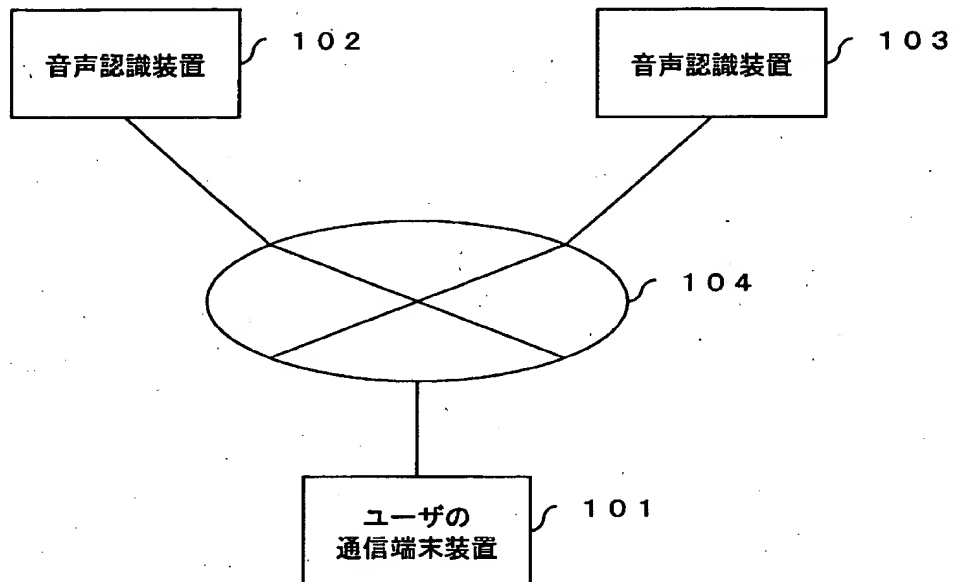
本実施の形態に係る通信端末装置が入力音声を音声認識する手順を説明するフローチャートである。

【符号の説明】

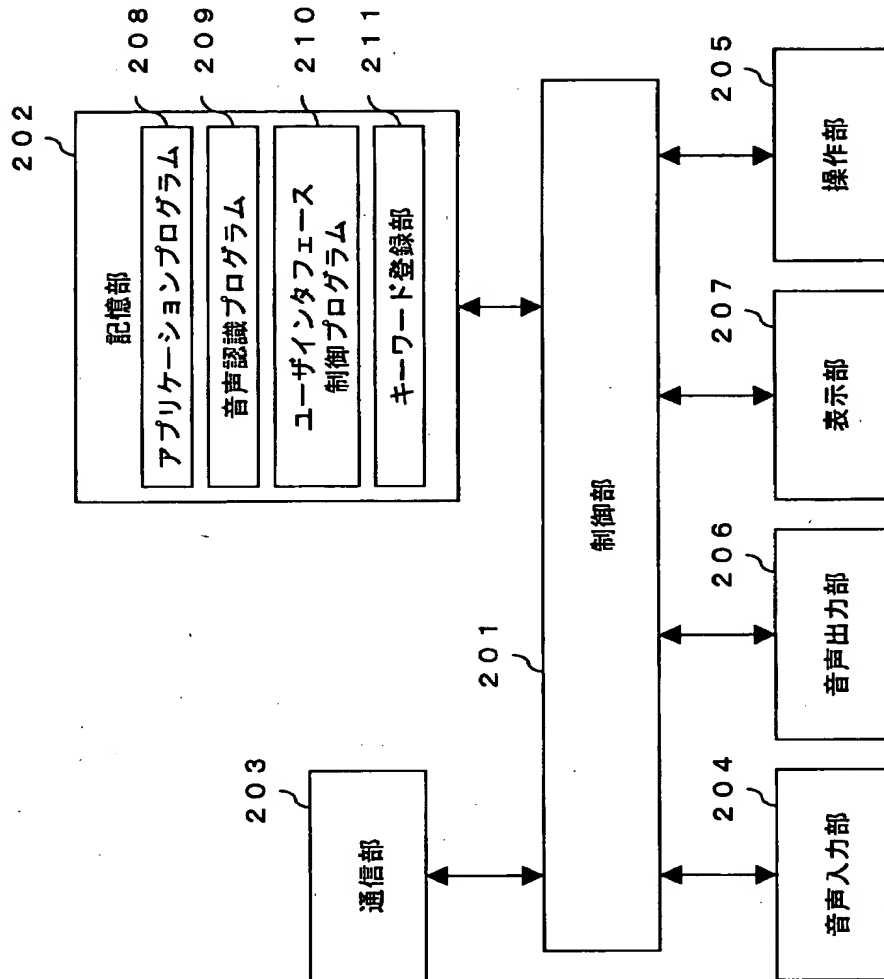
- 1 0 1 通信端末装置
- 1 0 2, 1 0 3 音声認識装置
- 1 0 4 ネットワーク
- 2 0 1 制御部
- 2 0 2 記憶部
- 2 0 3 通信部
- 2 0 4 音声入力部
- 2 0 5 操作部
- 2 0 6 音声出力部
- 2 0 7 表示部
- 2 0 8 アプリケーションプログラム
- 2 0 9 音声認識プログラム
- 2 1 0 ユーザインタフェース制御プログラム
- 2 1 1 キーワード登録部

【書類名】 図面

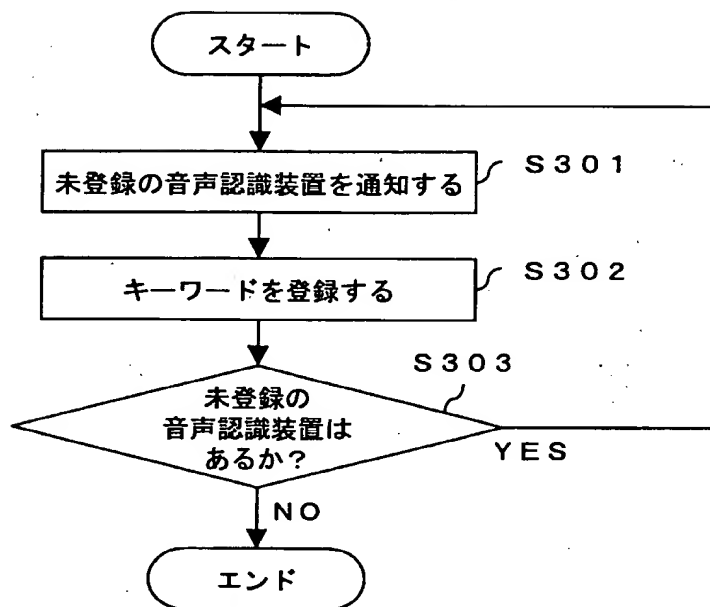
【図 1】



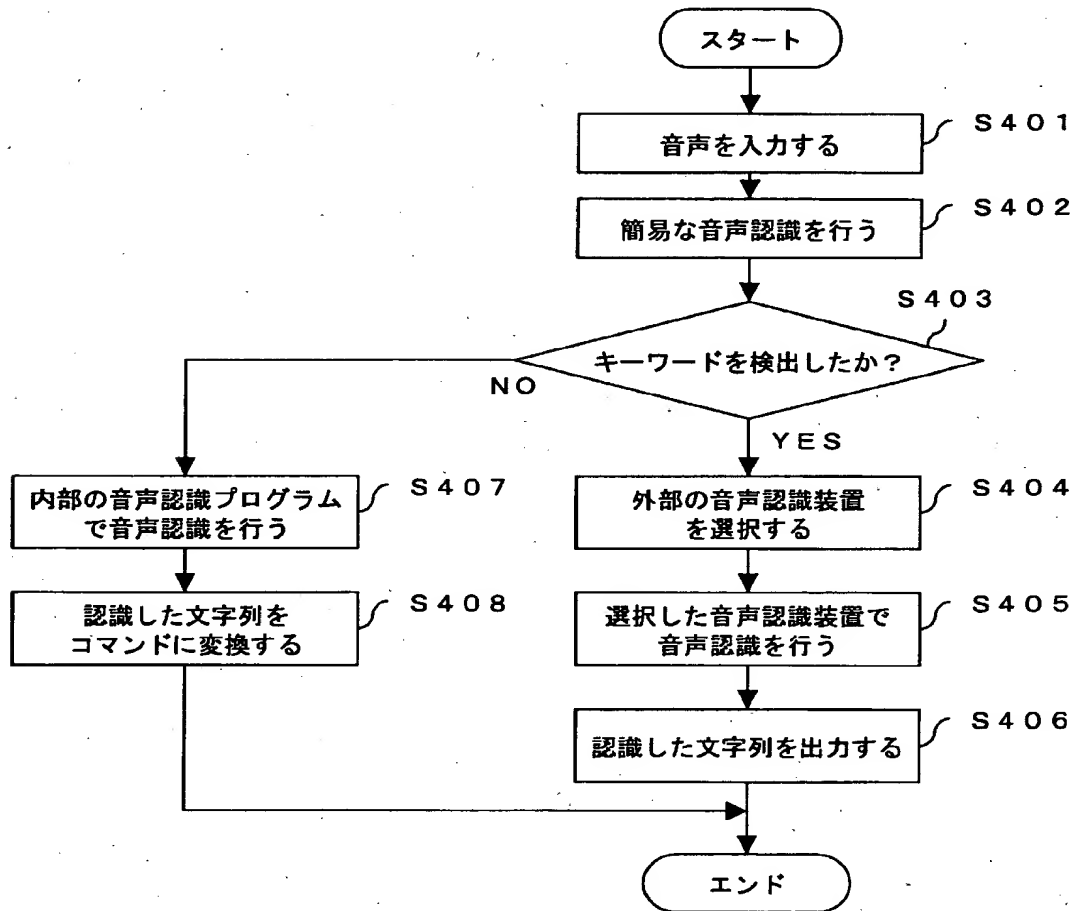
【図 2】



【図 3】



【図 4】



【書類名】 要約書

【要約】

【課題】 ネットワークに接続された複数の音声認識装置を利用して高精度な音声認識を効率的に行う。

【解決手段】 通信端末装置 1 0 1 は、ネットワーク 1 0 4 に接続された高精度な音声認識装置 1 0 2 又は 1 0 3 を指定する指定情報をユーザが発声した音声の中から検出する。指定情報が検出された場合には、検出された指定情報に対応する音声認識装置 1 0 2 又は 1 0 3 を用いてユーザが発声した音声を音声認識する。

【選択図】 図 1

出 願 人 履 歴 情 報

識別番号

[000001007]

| | |
|----------|-------------------|
| 1. 変更年月日 | 1990年 8月30日 |
| [変更理由] | 新規登録 |
| 住 所 | 東京都大田区下丸子3丁目30番2号 |
| 氏 名 | キャノン株式会社 |